関係データベースデザインの数学的基礎理論について (情報科学の数学的基礎理論と応用)

増永, 良文

数理解析研究所講究録
353巻 1979年 30-39

関係データベースデザインの数学的基礎理論について

東北大学 電気通信研究所
増永 良文

1. はじめに

関係データベースを論理設計する手法は大別して、1)分解的手法 2)合成的手法の二つがある。 1)の分解法は Codd[1] の導入した関数従属性, Fagin[2] の導入した多値従属性等の情報を使い、与えられた初期関係スキーマを通常第三正規形, Boyce-Codd 正規形, あるいは第四正規形である基本関係スキーマと呼ばれる関係スキーマの集合に情報無損失分解してゆく手法である。 勿論、分解の目的は良く知られている様に、正規形で云わゆる storage anomalies を少なくすることが出来るからである。

さて、本稿の目的は最初既に示されている幾つかの分解的手法に則ったデザイン法を概括し、次いでこれら分解法の共通の欠点である情報無損失分解が下向き(与えられた関係スキーマを分解していく丈)の過程であるということを改善す

べく，著者が既に導入している関係間関数従属性なる概念を紹介し，この概念が従来の分解法より更に強力なデザイン法のデザイン tool となっていることを示す。

2．基礎的事項

(a)．関係スキーマは関係の構造的，意味論的枠組を与えるものであり，関係データベースの設計とは関係スキーマの組を設計することである。実際のデータはこの枠組がとる値であって，これを関係スキーマのインスタンスという。
$R=(\alpha, \mu, F)$，$\alpha=\{A_1, \cdots, A_n\}$は属性集合，$\mu$はドメイン写像，$\mu(A_i)=D_i$（$D_i$はドメイン），$F$は以下で述べる関数従属性の集合，は関係スキーマである。
Rのインスタンスrは$D_1\times\cdots\times D_n$の有限部分集合で，$F$の各元と両立するものを云う。

(b) Fの元は$\beta, \gamma \subseteq \alpha$，として$\beta \rightarrow \gamma$なる形の陳述である。$\beta\rightarrow\gamma$と$r$が両立するとは次の条件が成立する時を云う。$\forall t, t' \in r,\ t_\beta = t'_\beta \Rightarrow t_\gamma = t'_\gamma$ 。
ここにt_βはタップルtのβ-部分である。

(c) $R(\alpha)$を関係スキーマ，$\beta\rightarrow\gamma$をRの関数従属性とする時．$R_1(\beta\cup\gamma)=\pi_{\{\beta\cup\gamma\}}(R)$，$R_2(\alpha\setminus\gamma)=\pi_{\{\alpha\setminus\gamma\}}(R)$，ここに$\pi$は射影演算，と定義すると，$R$と$R_1$と$R_2$

に分解するこの分解の仕方は情報無損失である。ここに情報無損失の定義はRの任意のインスタンスrに対して $r = \pi_{\{\beta\cup\gamma\}}(r) * \pi_{\{\alpha\cup\gamma\}}(r)$ が成立することである。*は結合演算を表わす。

(d) 関係スキーマRは全ての Non-prime 属性(どの candidate key にも属さない属性)が各 candidate key に全関数従属の時に第二正規形という。関係スキーマ R(SUPPLIER, CITY, POPULATION), SUPPLIER → CITY, CITY → POPULATION, は第二正規形にある。第二正規形であってかつ全ての non-prime 属性は各 candidate key に非遷移的に従属している時、第三正規形であるという。Boyce-Codd 正規形とは全ての全関数従属性の決定子が candidate key になる時をいう。第三正規形であって Boyce-Codd 正規形でない有名な例は関係スキーマ R(STUDENT, TEACHER, COURSE), {STUDENT, COURSE} → TEACHER, TEACHER → COURSE, である。これは第三正規形であるが TEACHER が key でないので Boyce-Codd 正規形でない。Rが第四正規形であるとは多値従属性 $\beta \twoheadrightarrow \gamma$ が存在すれば、Rの全属性が β に関数従属の時をいう。Boyce-Codd 正規形であって第四正規形でない例は R(COURSE, TEACHER, TEXT), COURSE →→ TEACHER, COURSE →→ TEXT がある[4]。

(e) 基本関係が才三，Boyce-Codd，才四正規形である関係スキーマの分解法を才三，Boyce-Codd，才四正規形分解法と云う。

3. 才三正規形分解法の能力的限界.

以下例題で示す様に才三正規形分解法による設計法は初期デザインをデザイナーがどの様に与えるかに強く依存してしまう。これはBoyce-Codd，才四正規形いずれの分解法にも共通した欠点であり，次に示す様に分解法がtop-downの性格を有するということから来る能力的限界による。

[例題] 科学研究所のデータベースの設計例：研究所には研究者，タスク，プログラム，予算の概念があるとする。次の二つの初期デザインを与える。

初期デザインA：$R_{研-タ}$(SNAME, T.NAME, T.LEADER, T.BUDGET), S.NAME → T.NAME, T.NAME → T.LEADER, T.NAME → T.BUDGET, $R_{研-プ}$(S.NAME, P.NAME, P.LEADER, P.BUDGET), S.NAME → P.NAME, P.NAME → P.LEADER, P.NAME → P.BUDGET，からなる。

初期デザインB：$R_{研-タ}$，$R_{タ-プ}$(T.NAME, P.NAME, P.LEADER, P.BUDGET), T.NAME → P.NAME, P.NAME → P.LEADER, P.NAME → P.BUDGET，からなる。

これら初期デザインに定義されている関数従属性を使って才

三正規形分解を行なうと次の最終デザインをうる。最終デザインA，Bは各々初期デザインA，Bから得られたもの。

<u>最終デザインA</u>： R_{st}(S.NAME, T.NAME), R_t(T.NAME, T.LEADER, T.BUDGET), R_{sp}(S.NAME. P.NAME), R_p(P.NAME, P.LEADER, P.BUDGET).

<u>最終デザインB</u>： R_{st}, R_t, R_{tp}(T.NAME, P.NAME), R_p.

問題点は最終デザインAは初期デザインBから，最終デザインBは初期デザインAからはどの様な分解をしても決して得られないということである。このことからデザイナーは初期デザインの選定に苦悩するかもしれない。これが従来の分解法の欠点であり，よりフレクシブルなデザイン法が要求される原因でもある。この問題点を次に関係間関数従属性の概念を入れて打破する。

4. 関係間関数従属性の導入と応用

$R_1=(\alpha_1, \mu_1, F_1)$, $R_2=(\alpha_2, \mu_2, F_2)$ が結合可能とは次の二つの条件が成立する時を云う。

(1) $\alpha_1 \cap \alpha_2 \neq \phi$　(2) $\mu_1 | \alpha_1 \cap \alpha_2 = \mu_2 | \alpha_1 \cap \alpha_2$.

更に (3) $\beta \to \gamma$, $\beta \cup \gamma \subseteq \alpha_1 \cap \alpha_2$, が共に F_1 と F_2 の元であるならこれら二つの関数従属性は時変関数の意味で一致する，という条件を関数従属性の一意性の保証の為課する。

そして改めてR_1とR_2は結合可能と云う。注意することはこの結合の概念は極めて一般的であることである。その意味は本章終りに記す。$R_1 * R_2$でR_1とR_2の結合を表わす。$R_1 * R_2 = (\alpha_1 \cup \alpha_2, \mu_{12}, F_1 \cup F_2)$, $\mu_{12} | \alpha_1 = \mu_1$, $\mu_{12} | \alpha_2 = \mu_2$である。

次にR_1とR_2を結合可能な関係スキーマとし，$\beta, \gamma \subseteq \alpha_1 \cup \alpha_2$，しかし$\beta, \gamma \subseteq \alpha_1$でも$\beta, \gamma \subseteq \alpha_2$でもないとする。この時$R_1$と$R_2$の間の関係間関数従属性（以下IFDと略記）は$\beta \to \gamma$なる形の陳述である。この意味は組$(R_1, R_2)$の任意の時刻$t$におけるインスタンスの組$(r_1, r_2)$に対して，もし$r_1 * r_2 \neq \phi$ならば，$\forall u, u' \in r_1 * r_2$に対して$u_\beta = u'_\beta \Rightarrow u_\gamma = u'_\gamma$が成立することである。IFDsは$R_1$と$R_2$の間に任意に定義されるものではなく，次の命題群で規定される制約を有する。

命題1. r_1, r_2を結合可能な二つの関係スキーマR_1とR_2のインスタンスとし$r_1 * r_2 \neq \phi$とする。もしIFD $\beta \to \gamma$（以下一般性を失なうことなく$\gamma \subseteq \alpha_2$と仮定）が$R_1$と$R_2$の間に存在するならば，$(\alpha_1 \cup \beta) \cap \alpha_2 \to \gamma$は$\pi_{\alpha_2}(r_1 * r_2)$に対して成立しなければならない関数従属性である。

命題2. 上述と同じ条件のもとに次が成立する。

(i) もし$\alpha_2 \cap \beta \neq \phi$でかつ$(\alpha_2 \cap \beta) \to \gamma$が$\pi_{\alpha_2}(r_1 * r_2)$

に対して成立しないならば，$\pi_{(\alpha_1 \cap \beta)}(r_1 * t_{(\alpha_1 \cap \alpha_2)}) \cap \pi_{(\alpha_1 \cap \beta)}(r_1 * t'_{(\alpha_1 \cap \alpha_2)}) = \phi$ が $t_{(\alpha_2 \cap \beta)} = t'_{(\alpha_2 \cap \beta)}$ で $t_\gamma \neq t'_\gamma$ なる条件を満足する様な $\pi_{\alpha_2}(r_1 * r_2)$ の全タップル t と t' に対して成立する。

(ii) もし $\alpha_2 \cap \beta = \phi$ ならば，$\pi_\beta(r_1 * t_{(\alpha_1 \cap \alpha_2)}) \cap \pi_\beta(r_1 * t'_{(\alpha_1 \cap \alpha_2)}) = \phi$ が $t_\gamma \neq t'_\gamma$ なる条件をみたす $\pi_{\alpha_2}(r_1 * r_2)$ の全タップル t と t' に対して成立する。

__命題3.__ r_1 と r_2 を結合可能な二つの関係スキーマ R_1 と R_2 のインスタンスとし $r_1 * r_2 \neq \phi$ とする。 β と γ を $\alpha_1 \cup \alpha_2$ の部分集合で $\gamma \subseteq \alpha_2$ しかし $\beta \subseteq \alpha_2$ ではないとする。 この時，もし命題1及び2の条件が成立するならば $\beta \rightarrow \gamma$ は対 (r_1, r_2) に対して成立するIFDである。

__定理1.__ R_1 と R_2 を結合可能な二つの関係スキーマとする。β, γ を $\gamma \subseteq \alpha_2$ ではあるが $\beta \subseteq \alpha_2$ ではない $\alpha_1 \cup \alpha_2$ の部分集合とする。 この時，$\beta \rightarrow \gamma$ なる形の陳述が対 (r_1, r_2)，ここに r_1 と r_2 は R_1 と R_2 のインスタンスで $r_1 * r_2 \neq \phi$)，に対して成立する為のIFDである為の必要かつ十分条件は命題1及び2の条件が成立することである。

__定理2.__ 定理1と同様の条件のもとで，$\beta \rightarrow \gamma$ なる形の陳述が R_1 と R_2 の間のIFDである為の必要かつ十分条件は $\beta \rightarrow \gamma$ が $r_1 * r_2 \neq \phi$ なる R_1 と R_2 の全てのインスタンスの組

(r_1, r_2) に対して成立する IFD であることである。

さて，上述の結果のデータベースデザインへの応用を以下考察する。まず R_1 と R_2 を結合可能な関係スキーマとし全ての時刻 t に対して，そのインスタンスの組 (r_1, r_2) が $\pi_{(\alpha_1 \cap \alpha_2)}(r_1) = \pi_{(\alpha_1 \cap \alpha_2)}(r_2)$ を満す時，R_1 と R_2 は情報無損失結合可能と云うことにする。

定理3. 情報無損失結合可能な二つの関係スキーマ R_1 と R_2 の間に IFD $\beta \to \gamma$ $(\gamma \subseteq \alpha_2)$ が存在したとする。この時 R_1 と R_2 は情報を失なうことなく次の三つの関係スキーマに置き換えられる。$R'(\beta \cup \gamma)$，$R''(\alpha_1 \cup (\beta \cap \alpha_2))$ と $R'''(\alpha_2 \setminus \gamma)$。

[略証] IFD $\beta \to \gamma$ により存在を保証された R_2 の関数従属性 $((\alpha_1 \cup \beta) \cap \alpha_2) \to \gamma$ を使い R_2 を $R_{21} = \pi_{(((\alpha_1 \cup \beta) \cap \alpha_2) \cup \gamma)}(R_2)$ と $R_{22} = \pi_{(\alpha_2 \setminus \gamma)}(R_2)$ に分解する。この時 R_1 と R_{21} は情報無損失結合可能で，かつ $\beta \to \gamma$ は R_1 と R_{21} 間の IFD となるので R_1 と R_{21} は情報を失なうことなく結合され $R_1 * R_{21}$ を構成し，$\beta \to \gamma$ はこの関数従属性となる。そこで $R_1 * R_{21}$ を $\beta \to \gamma$ で情報無損失分解すると $\pi_{(\beta \cup \gamma)}(R_1 * R_{21})$，$\pi_{(\alpha_1 \cup (\beta \cap \alpha_2))}(R_1 * R_{21})$ になる。☒

本結果は IFD を使ったデータベースのデザイン法を論ずる場合本質的である。つまり 定理3の条件のもとに，最

前に R_2 を $((\alpha_1 \cup \beta) \cap \alpha_2) \to \gamma$ で分解せずに，$R_1 * R_2$ を作ってしまうと，その後 $R_1 * R_2$ の関数従属性として定義された $\beta \to \gamma$，あるいは $((\alpha_1 \cup \beta) \cap \alpha_2) \to \gamma$ をどの様に使っても定理3中の R', R'', R''' の組は得られない。これは初期デザインとして $R_1(\alpha_1)$，$R_2(\alpha_2)$，IFD $\beta \to \gamma$ なる構成を与えるのと $R(\alpha_1 \cup \alpha_2)$，$\beta \to \gamma$，$((\alpha_1 \cup \beta) \cap \alpha_2) \to \gamma$ なる構成を与えるのとの根本的差異である。又後者の構成から導出される最終デザインは全て前者の構成からも導出出来かつ上述のことから，IFDs を導入した関係データベースのデザイン法は，それを導入しないデザイン法に比べ真にデザイン能力に於いて強力であることが判る。第三正規形分解法であって，初期デザインとして複数個の関係スキーマと IFDs を与えるデザイン法を拡張された第三正規形分解法と呼ぶことにする。この分解法では容易に確かめられる様に，第三章の例題を使えば，最終デザイン A を初期デザイン B から，又最終デザイン B を初期デザイン A から得ることが出来，拡張された第三正規形分解法は真にデザイン能力において，従来の第三正規形分解法より強力な手法であることが検証される。

5. 結論

関係データベースデザインの分解的手法が概括され、それらに共通な欠点である下向きの過程がデザイン能力を限定していることを指摘した。この欠点を改善すべく、関係間関数従属性なる新しい概念が定義導入された。この結果得られた拡張された第三正規形分解デザイン法は従来の第三正規形分解法より真にデザイン能力で強力な手法であることが示された。

[文献]

[1] Codd, E.F. Further normalization of the data base relational model. Courant Comp. Sci. Symp. (1971)

[2] Fagin, R. Multivalued dependencies and a new normal form for relational databases, TODS 2.3 (1977)

[3] Fagin, R. The decomposition versus the synthetic approach to relational database design. 3rd VLDB Conf. Rec. (1977)

[4] Date, C.J. An Introduction to database systems. 2nd edition, Addison-Wesley (1977)

以上.